2020 Census
Program Management Review

Modeling and Analysis for the Cost-Quality Trade-Off

Pat Cantwell

Decennial Directorate

April 8, 2015

United States Census Bureau | U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Final v1.0

Topics

- Evaluating cost and quality
- The Microsimulator

Evaluating Cost and Quality

- Ideal
  - Want a simple measure of cost and quality for entire operation
  - Want to consider “downstream” effects
- Pragmatic
  - Not so simple; no one method
  - Cost and quality are difficult to measure
  - Must identify appropriate metrics

Measuring Quality

- What to measure, and level of aggregation:
  - Geographic: nation, state, others
  - Demographic: race, Hispanic origin, age groups
- How to measure: sums of errors, absolute errors, weighted errors, . . .
- Standard of comparison, what to use as a baseline

Measuring Cost

- Compatible with Census Bureau's cost model, cost estimates
- Rough measures
  - For example, NRFU: workload, weighted; number of visits
- Combining across operations, common measures
  - Dollars
  - Other measures for cost model, e.g., productivity

Evaluating Cost and Quality, Summary

Determining appropriate metrics:

- Narrowing the list
- Which ones answer the proper questions?
- Which ones can we obtain accurately?

Two approaches to evaluating cost and quality:

- Analyze components one at a time
- Analyze from beginning to end

The Microsimulator

What is a Microsimulator?

- "Simulator": Computer program that runs scenarios or options as specified, while introducing realistic randomness
- "Micro": Starts at level of address, housing unit, person
- Allows us to run through scenarios many times, to see a **distribution of possible results**
- Input: Statistical models, data
- Output: Data, results, aggregates
  - Quality metrics: quality of census response, of census count
  - Cost measures: workload, visits; miles driven, productivity

Primary Purpose of the Microsimulator

- Provides information on the trade-off between cost and quality; use it as we shape the design decision, the operational plan for the 2020 Census

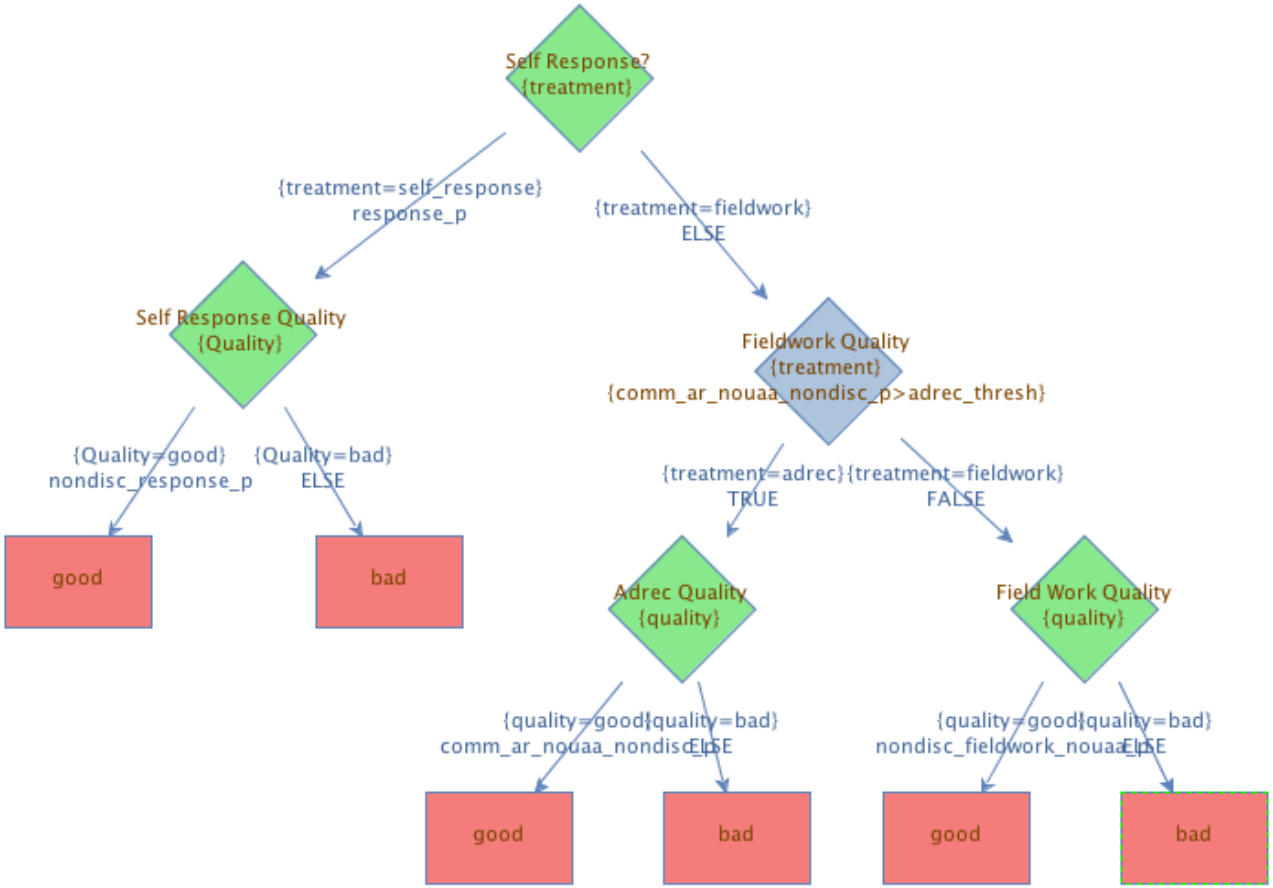
- Replaces our prior efforts on life cycle analysis, by allowing us to:
  - combine various activities into options
  - evaluate the trade-offs

Benefits of a Microsimulator

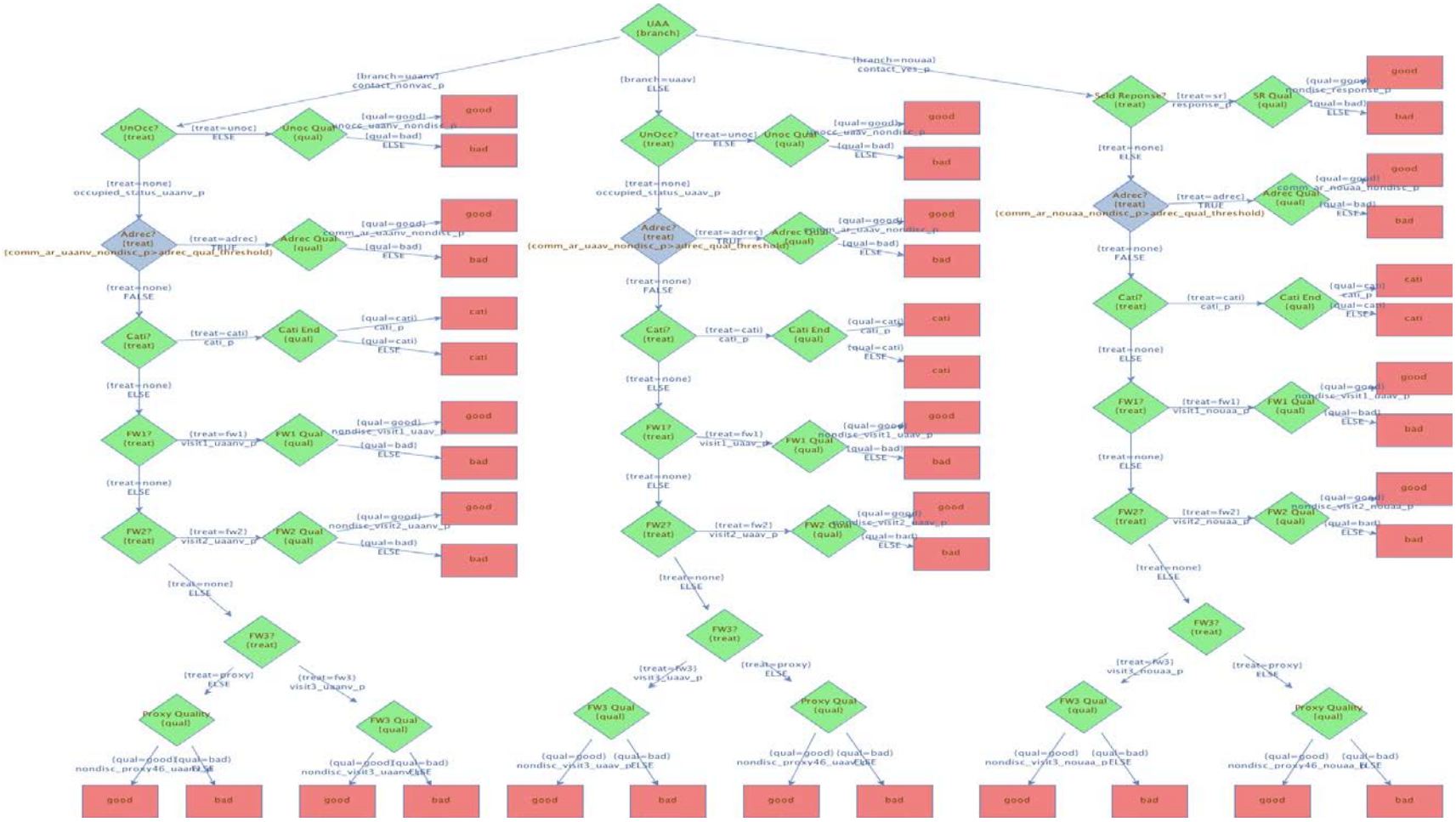
Enables us to . . .

- Simulate individual actions or decisions, then track the result through a *series* of actions

Simulating a Series of Actions



Simulating a Series of Actions (cont.)



Benefits of a Microsimulator

Enables us to . . .

- Simulate individual actions or decisions, then track the result through a *series* of actions
- Combine a string of activities into one complete "design option" or scenario

Example of Activities

Address Canvassing

- Option 1: canvass in field 100%
- Option 2: canvass in field 40%
- Option 3: canvass in field 20%

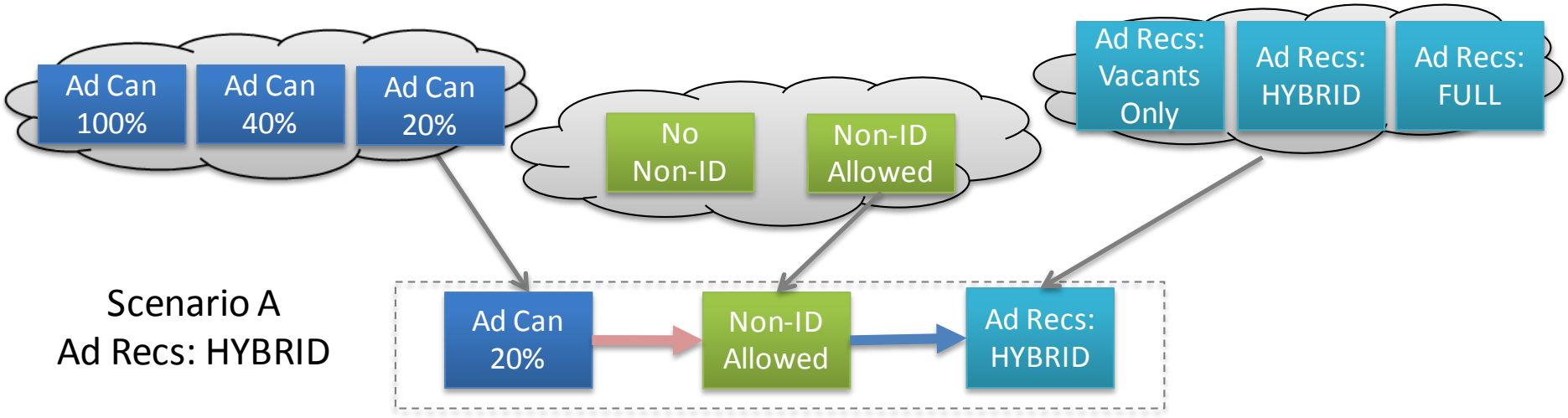
Self-Response

- Option 1: internet, mail, non-ID not allowed
- Option 2: internet, mail, non-ID allowed

Use of Administrative Records

- Option 1: remove vacants (and deletes) (VACANTS ONLY)
- Option 2: remove vacants, use Ad Recs after 1 NRFU visit (HYBRID)
- Option 3: remove vacants, use Ad Recs for all nonrespondents (FULL)

Building a Design Option or Scenario

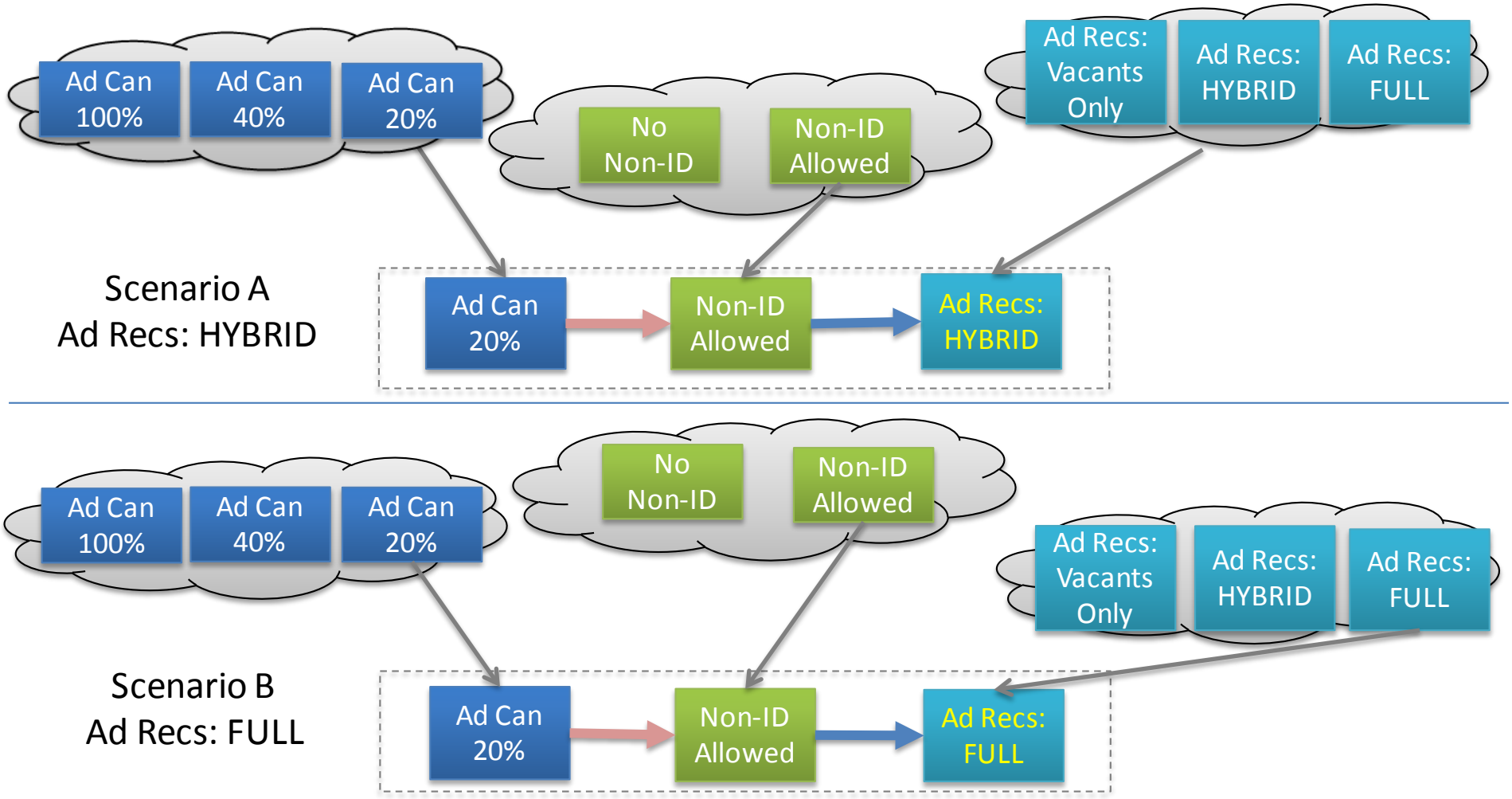


Benefits of a Microsimulator

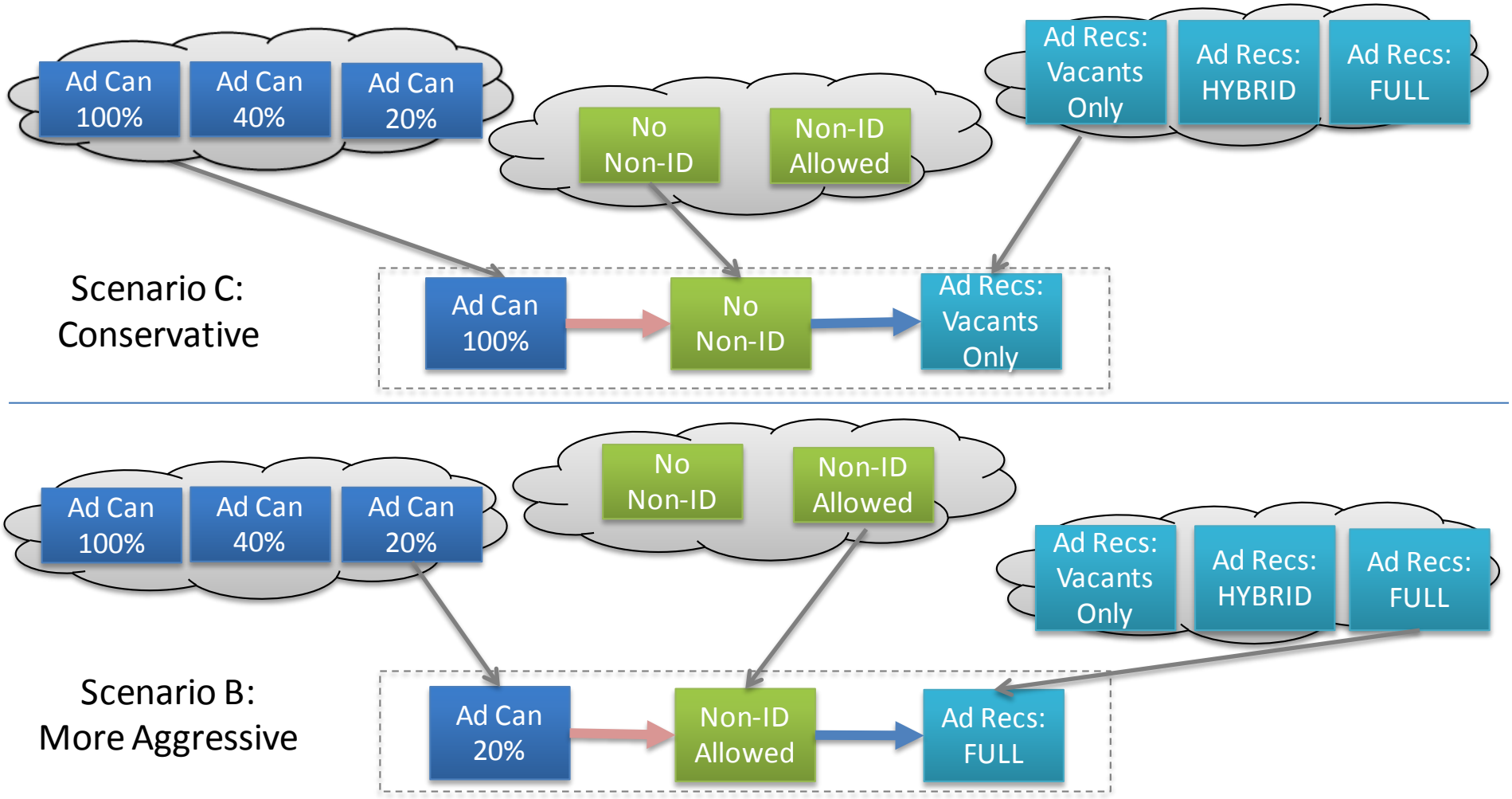
Enables us to . . .

- Simulate individual actions or decisions, then track the result through a *series* of actions
- Combine a string of activities into one complete "design option" or scenario
- Compare two or more scenarios

Analyzing Components One at a Time: Use of Ad Recs



Analyzing from Beginning to End



Benefits of a Microsimulator

Enables us to . . .

- Simulate individual actions or decisions, then track the result through a *series* of actions
- Combine a string of activities into one complete "design option" or scenario
- Compare two or more scenarios
- Observe trade-off between, say, cost and quality
- Capture randomness at housing-unit level; see how this produces a range of potential aggregate results

Comparing Cost and Quality

(each dot represents an iteration of the simulation)

Limitations of a Microsimulator

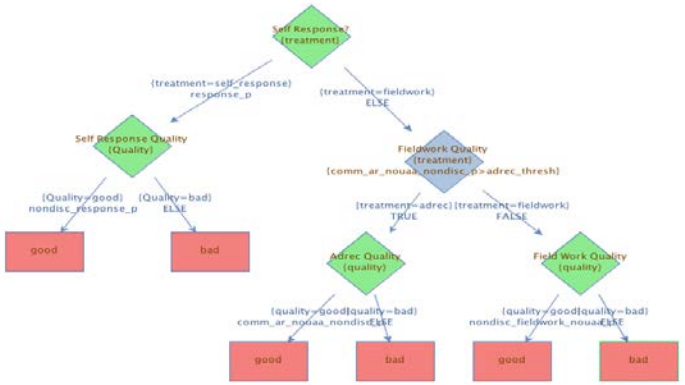
- Results are only as good as the models and input data
- Results rely on validity of assumptions
  - variation, randomness introduced
  - relationships between activities
- Developing, testing, and programming the underlying models: each is time intensive
- Computing time to run the simulation (many iterations) can be quite long

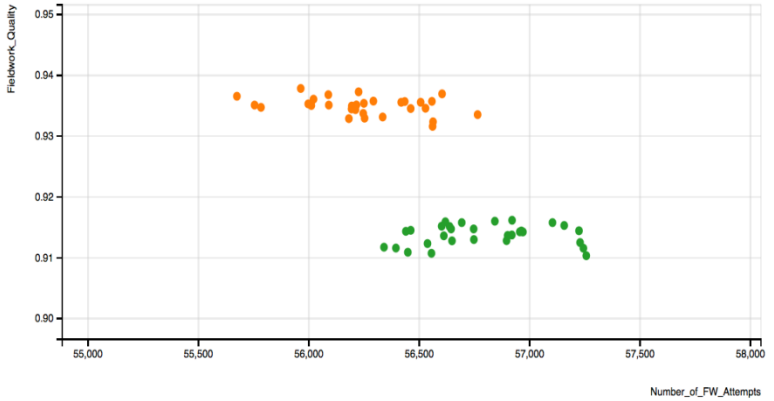
Benefits of This Microsimulator

1. Specifying and building scenarios graphically

2. Changing input parameters

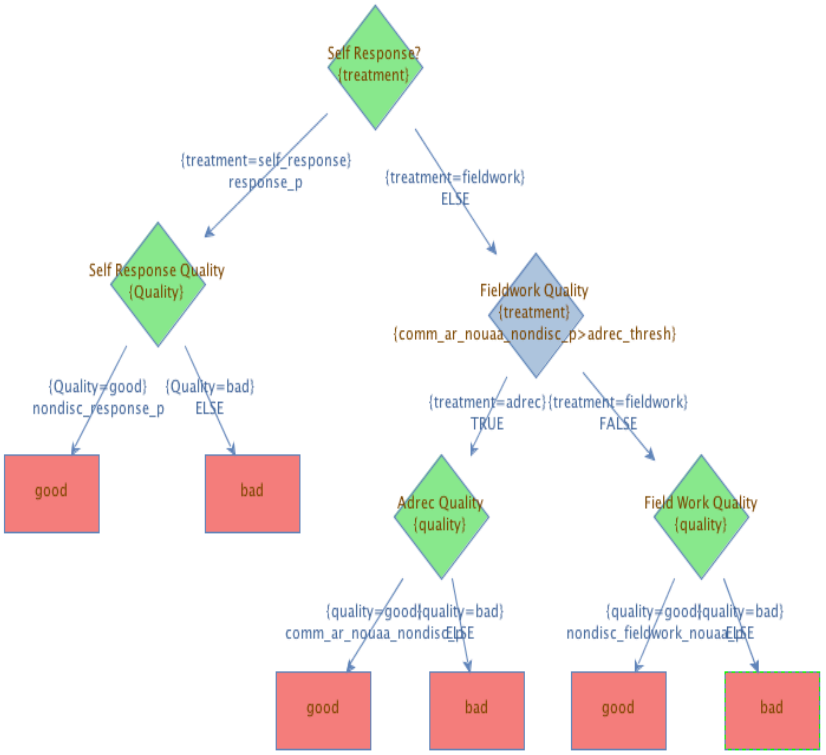
3. Visually exploring output data





1. Building Scenarios Graphically

- Generate logic diagrams, graphically specify scenarios
- Manipulate the graph; underlying Python code is generated *automatically*
  - simpler to build and change
  - increases coding accuracy



2. Changing Input Parameters

- For each unit (e.g., address), models produce, say,

  Pr{ unit responds via internet },

  Pr{ unit has "good" administrative records }, etc.

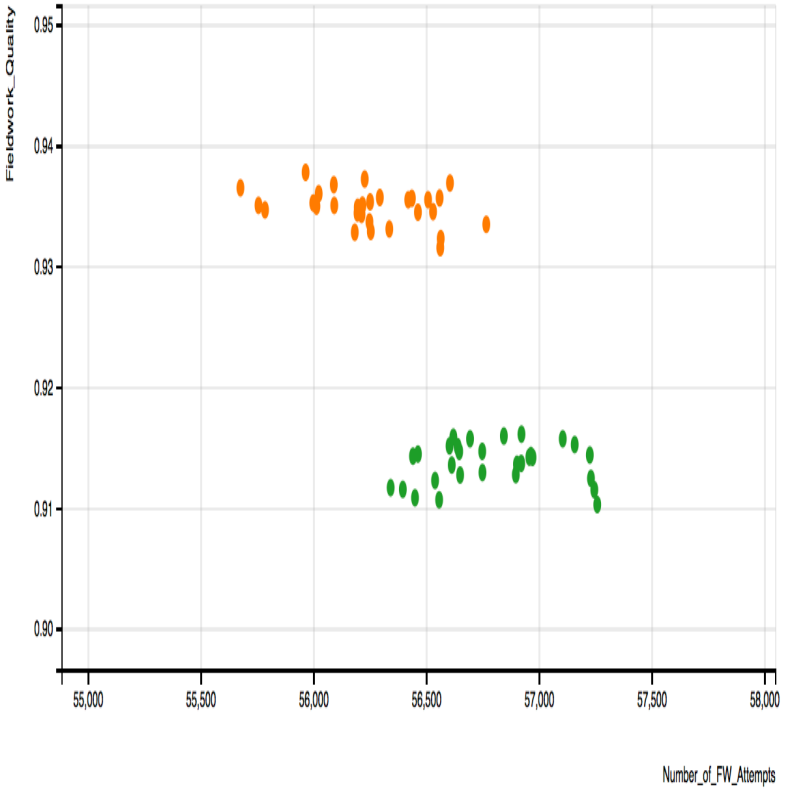
- Easy to change input parameter to, say, Pr{ … } + Δ
  - for all units
  - for a specified subset of units

- Allows us to easily compare results under differing inputs

3. Exploring the Output Data

Easy to explore results visually; change domains, metrics

- Aggregates: nation, states, etc.
- Demographic or other domains
  - race or origin of householder
  - used vs. didn't use admin records
- Cost metric
  - number of NRFU cases
  - total number of NRFU visits



Current State of Microsimulator

- Developed two major activities:
  - self-response options
  - use (or not) of administrative records
- Currently developing models and gathering data for address canvassing options
- Simulation programmed in two languages (Python and SAS), for verification and risk mitigation

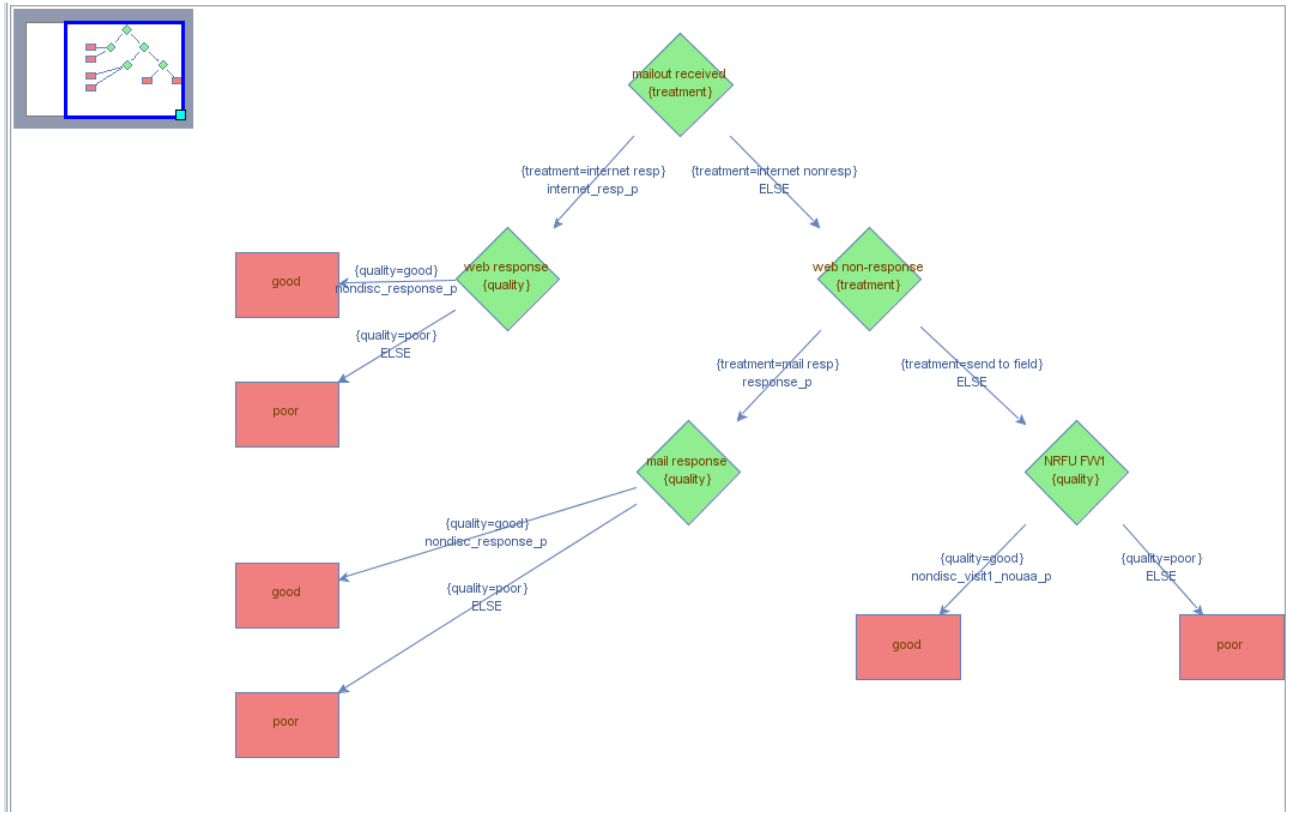
Broader Application of Microsimulator

- FY15: Produce trade-off data for design decision, operational plan for the 2020 Census
- Beyond FY15: Refine models; add other census operations to the sequence; evaluate options
- Eventually: Share software with other parts of the Census Bureau; encourage adoption of simulation techniques and use of statistical models

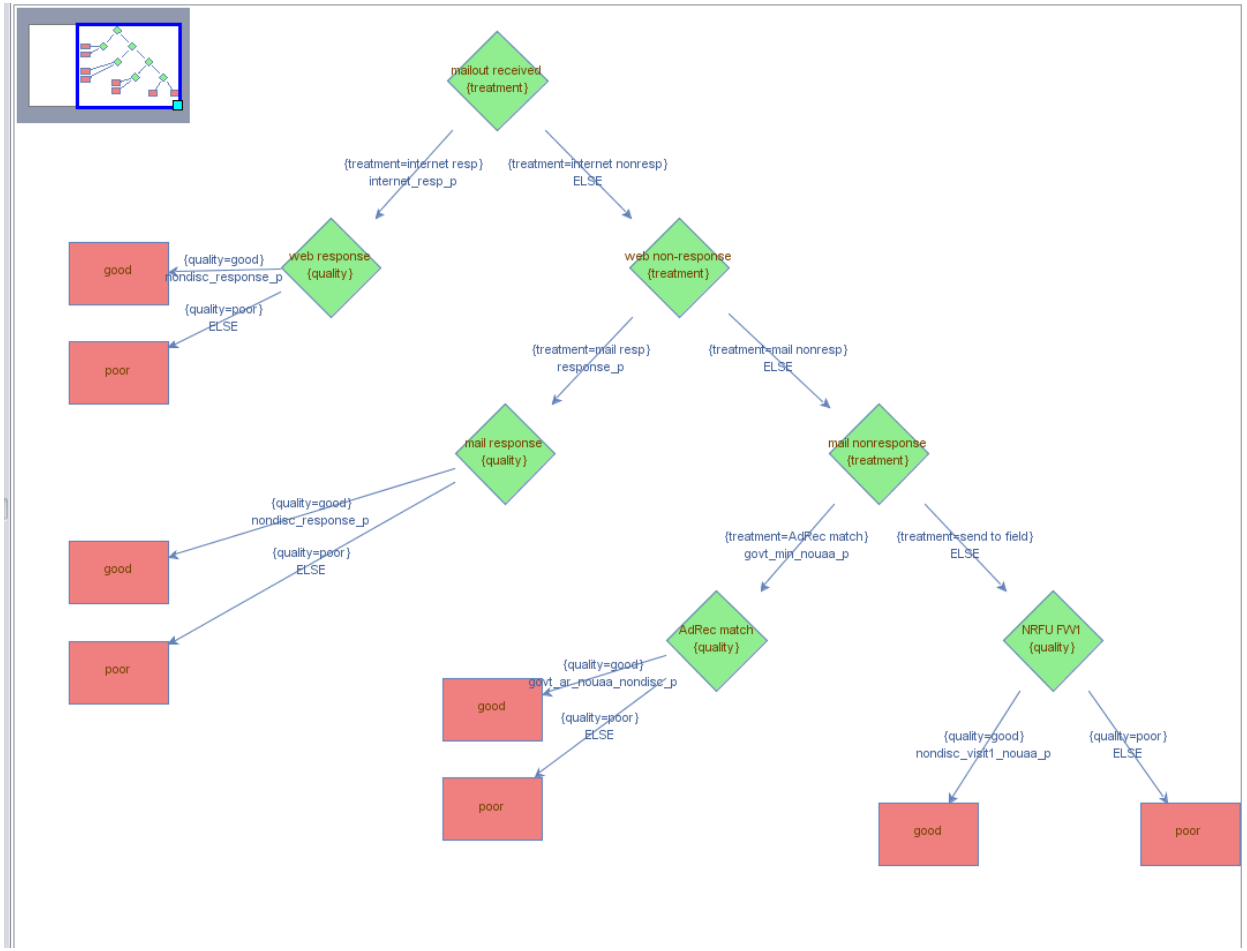
Questions

census.2020.program.management.review@census.gov

Building Scenarios Graphically (cont.)

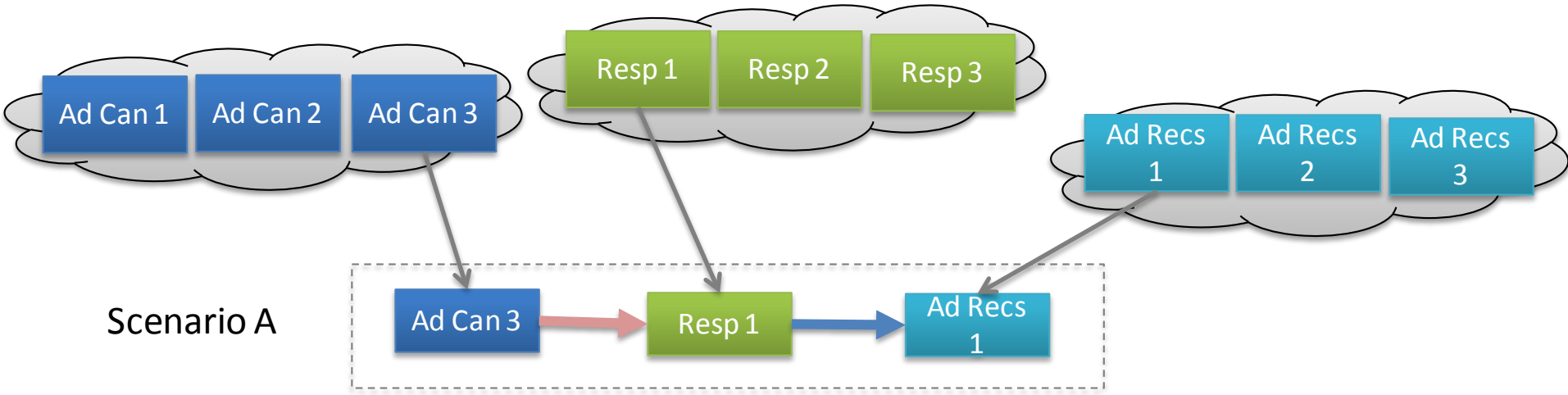


Building Scenarios Graphically (cont.)



Combining Activities

- Start with options for several activities; for example,
  - address canvassing
  - self-response
  - use of administrative records (or not)
- Select one option for each activity
- Combine the individual options into one “design option” or “scenario”



Range of Resulting Values, Randomness

- The simulation produces "point clouds" showing the range of possible outcomes
- Which scenario works better on average?
- But how much uncertainty do Scenarios A, B incur?